

Disability testing and retirement¹

Helmuth Cremer, Jean-Marie Lozachmeur and Pierre Pestieau

March 2004, revised November 2004

¹An earlier version of this paper has been circulated under the title “Disability, preferences for leisure and retirement decision”.

Abstract

This paper studies the design of retirement and disability policies. It illustrates the often observed exit from the labor force of healthy workers through disability insurance schemes. Two types of individuals, disabled and leisure-prone ones, have the same disutility for labor and cannot be distinguished. However, they are not counted in the same way in social welfare. Benefits depend on retirement age and on the (reported) health status. We determine first- and second-best optimal benefit levels and retirement ages and focus on the distortions which may be induced in the individuals' retirement decision. Then we introduce the possibility of testing which sorts out disabled workers from healthy but retirement-prone workers. We show that such testing can increase both social welfare and the rate of participation of elderly workers; in addition disabled workers are better taken care of. It is not optimal to test all applicants, nor to apply testing to all types of benefits. Surprisingly, the (second-best) solution *may* imply later retirement for the disabled than for the leisure prone. In that case, the disabled are compensated by higher benefits.

1 Introduction

Over the last decade the future of pension systems has been a widely debated issue. The challenges posed by the ongoing process of demographic aging are by now understood. However, aging is only part of the story. The generalized decrease of entry ages into retirement is also a significant problem. There are a number of explanations for this trend. The dominant one is the implicit taxation on continued activity of elderly workers. This argument is based on the idea that workers have, individually or collectively, a large degree of discretion concerning retirement age. In that line, Gruber and Wise (1999, 2004) have shown for a dozen of OECD countries that households often face very powerful financial incentives pushing them to retire early.

Some of these incentives may result from the deliberate choice of decision makers: early retirement schemes in some declining industries, disability insurance for disabled workers, unemployment insurance for workers who have no chance to find another job. It might be the case that policy makers have used too much of these policies with the complicity of unions and management. On some occasions programs might also have been diverted from their original purpose. For instance, unemployment and disability insurance have at times been used not only to compensate for “genuine” unemployment or disability but also to allow healthy and still productive workers to withdraw from the labor force. When looking at the data on disability – for unemployment, the issue is trickier – it is clear that they do not only encompass disabled workers. They also include people who want to stop working and to benefit from a system with slacking control.

“Flaws” in control can be deliberate or not. They can certainly be due to bad policy design. However, it is not clear either that a systematic use of tight controls to all applicants for (say) disability benefits is necessarily the appropriate policy. The question is where to draw the line between disability benefits, which would be subject to more or less tight screening, and (early) retirement benefits which are conditional only on age (and on the individual’s retirement decision). This is one of the questions discussed in this paper. In particular, we show that without control a number of workers

who highly value leisure will use disability insurance as a way to leave the labor market. This has two consequences: a relatively low age of retirement but more importantly a relatively low level of disability benefits.

We use a simple model with three different types of individuals with identical productivity: hard working individuals, retirement-prone individuals and disabled individuals. At first sight, the latter two look very much alike. However, the disabled have a high disutility of labor because effort hurts them while the retirement-prone individuals have the same high disutility because they like leisure. The social planner would like to give more weight to the disabled than to the two other ones that we will call the hard-working and the lazy ones. However if it cannot sort out the lazy and the disabled, they will be given the same package (or at least the same utility level).

We then introduce an audit procedure which allows for such sorting out. More precisely, there is a testing procedure which reveals individuals health status so that the disabled can be identified. We show that such control can increase both social welfare and the rate of participation of elderly workers; in addition disabled workers are better taken care of. However, we also show that it is not optimal to test all applicants, nor to apply testing to all types of benefits. To some extent the solution without audit looks very much like what is observed in a number of countries where there is a clear abuse of disability insurance programs with the consequence of low participation rate of elderly workers and with an erosion of disability benefits for budgetary reasons. The solution with audit looks like what was prevailing at the start of social security programs and what a number of economists are now advocating to increase the age of retirement: stop to use disability insurance as an early retirement program.

We have already mentioned the papers which show that disability insurance and other programs are standard routes for exiting the labor force before the normal retirement age. In addition, there are papers that specifically deal with disability insurance. In the United States there is a lot of work on the interaction between disability insurance and the labor market. Parsons (1980) in a widely quoted paper shows that social security program explains a large fraction of the postwar decline in prime-age labor force participation. This study was followed by several ones which never disputed the disin-

centive effects of disability insurance but questioned their magnitude [see e.g. Bound (1989), Parsons (1980, 1991) and Gruber (2000)]. This literature shows that there are two issues: one concerns the number of truly disabled workers and the other concerns the number of beneficiaries from disability insurance. Putting aside employment policy considerations, the difference between these two numbers comes from the difficulty to sort out true disabled workers in a world of asymmetric information.

2 A simple model of disability insurance

2.1 Preferences and types of individuals

The setting is inspired by Cremer *et al.* (2004). Consider an individual with productivity w , dividing his lifetime with duration normalized to one into a period of full activity and a period of retirement. Assume away any liquidity constraint and posit zero rate of interest and time discount rate. The budget constraint is then given by:

$$\int_0^1 c(t) dt = \int_0^z w dt,$$

where $c(t)$ is instantaneous consumption and z ($z \leq 1$) the age of retirement. Lifetime utility is defined as

$$U = \int_0^1 u(c(t)) dt - \int_0^z r(t) dt,$$

where $u(\cdot)$ is a strictly concave instantaneous utility functions while $r(t)$ denotes an increasing function of effort disutility. As it will appear below $r(t)$ can be viewed as a mixture of taste for retirement leisure and of physical and mental inability to work further years. We assume that $r(t)$ varies across individuals and increases over lifetime. Under these assumptions, the utility function and the budget constraint can be rewritten as:

$$U = u(c) - R(z),$$

and

$$hc = wz,$$

where

$$R(z) = \int_0^z r(t)dt. \quad (1)$$

In other words, there is perfect smoothing of consumption over time. We shall use this reduced form for utility and budget constraint throughout the paper.

There are three types of individuals indexed $i = H, L, D$. We assume

$$r_H(t) < r_L(t) = r_D(t) \text{ for every } t \quad (2)$$

which from (1) also implies $R_H(z) < R_L(z) = R_D(z)$. Individuals of type H are hard-working or healthy; they have a low disutility of labor. Type D and L have the same (high) disutility of labor which is higher than that for type H . They thus have the same preferences over consumption and leisure. However, they differ in their “health status”. Specifically, we think of type D as truly disabled for which the high disutility of labor is due to a poor health. Individuals of type L , on the other hand, are in good health but are leisure prone (or for short “lazy”). Formally, one can thus think of the health status as a second dimension of heterogeneity. This heterogeneity is not reflected in the individuals preferences over consumption and length of activity. However, it determines the preferences of the social planner who roughly speaking will want to help the disabled but not the leisure prone individuals; see below for further discussion.

Observe that (2) can also be written as

$$\frac{R'_H(z)}{u'(c)} < \frac{R'_L(z)}{u'(c)} = \frac{R'_D(z)}{u'(c)}, \quad z \in [0, h], c > 0, \quad (3)$$

which can be interpreted as a single-crossing property. For any bundle in the (z, c) space, L and D type individuals have steeper indifference curves than H type individuals.

2.2 Policy

We now introduce the possibility of a disability/retirement insurance scheme consisting of a payroll tax τ and a (pension) benefit function $p(z)$ which depends on the retirement age. This tax-transfer scheme $(\tau, p(z))$ encompasses disability and retirement insurance programs. For simplicity we shall often refer to this as “social insurance”. It can be

represented by a non linear tax function

$$\begin{aligned} T(z) &\equiv \int_0^z w\tau \, dt - \int_z^1 p(z) \, dt \\ &= z\tau w - (1-z)p(z). \end{aligned}$$

Differentiating $T(z)$ with respect to z yields the marginal tax on postponed work that underlies the retirement insurance program

$$T'(z) = \tau w + p(z) - (1-z)p'(z).$$

This expression shows the cost of delaying retirement which is given by τw , the payroll tax plus $p(z)$, the foregone pension minus the increase in benefit ($p'(z)$), if any, over the length of retirement $(1-z)$. We refer to a social insurance program as “actuarially neutral” at the margin if $T'(z) = 0$.¹

For future reference, note that a consumer’s problem under a non-linear function $T(z)$ can be written as

$$\max_z \quad u(wz - T(z)) - R(z). \quad (4)$$

From the first-order condition we then obtain

$$\frac{R'(z)}{u'(c)} = w - T'(z). \quad (5)$$

In words, a non-zero marginal tax implies a wedge between the marginal rate of substitution between work and consumption and the marginal productivity of labor. This relationship is important for the interpretation of our results below.

3 Indistinguishable types with different social weights

Let us now return to the interpretation of $R(z)$, that is the disutility for prolonged activity. This disutility is affected both by taste and by health factors. Some people want to retire early because they have a strong preference for leisure and desire enjoy

¹We use the term “actuarial neutrality” (or fairness) regarding retirement. This has to be distinguished from fairness (marginal or average) regarding the return from saving, namely the relation between contribution and benefit. See on this Crawford and Lilien (1981) and Desmet and Jouten (2003).

a long retirement period. Some other people want to retire because physically they cannot work longer. To account for this apparent ambiguity, we distinguish disabled (D) and leisure-prone individuals (L). Clearly if disability were easily observable there would be no problem. There are however cases when it is only observable at some cost, the cost of medical tests.

Recall that D and L have the same function $R_i(z)$; in other words $R_L(z) = R_D(z)$. However the social planner is not ready to consider those two types as identical. We assume that the social planner wants to help the disabled workers but not the leisure-prone ones. More precisely, the high disutility for prolonged activity is considered as “legitimate” (and fully accounted for in welfare) for the disabled but not for the lazy individuals.² One way of introducing such a distinction is to assume a weighted social welfare function with lower weights on type L than on type D . This leaves open the issue of which weight to attach to H . To keep things simple, we introduce welfare weights φ_i ($i = H, L, D$) such that $\varphi_L < \varphi_D = \varphi_H$. A possible justification of this choice is to think that there is a majority of people who ex ante do not know whether they are H or D . They maximize their expected utility with some limited concern for a minority of leisure-prone workers who will never be disabled. An alternative approach with paternalistic rather than Paretian social preferences is considered in Section 4 where the disutility for the lazy type L is evaluated with R_H rather than R_L .

3.1 Full information optimum and laissez-faire

Denoting p_i the proportion of type i individuals, the first-best problem is given by:

$$\max_{c_i, z_i} \sum_i \varphi_i p_i [u(c_i) - R_i(z_i)] + \mu \sum_i p_i (wz_i - c_i). \quad (6)$$

This gives the efficiency conditions:

$$\frac{R'_H(z_H)}{u'(c_H)} = \frac{R'_L(z_L)}{u'(c_L)} = \frac{R'_L(z_D)}{u'(c_D)} = w, \quad (7)$$

²The assumption that $R_L(z) = R_D(z)$ is not important for our analysis. It is made for simplicity because it yields crisper results in some cases. The crucial feature of our model is that the two types are treated differently in the social welfare function. See also footnote 4

which, not surprisingly, are equivalent to (5) with $T'(z) = 0$. With objective (6) the social optimum is Pareto efficient and we have the usual first-best tradeoffs.

We also have:

$$\varphi_H u'(c_H) = \varphi_L u'(c_L) = \varphi_D u'(c_D) = \mu, \quad (8)$$

so that consumption levels and retirement ages are ranked as follows:

$$\begin{aligned} c_D &= c_H > c_L \\ z_D &< z_H \quad z_D < z_L \quad z_H \geq z_L. \end{aligned}$$

With this specification type D 's individuals end up retiring earlier than the two other ones. They consume as much as H and more than L . In other words they are compensated for their disability not only by retiring earlier but also by receiving higher benefits (which is conceivable in a social insurance setting).

The *laissez-faire* solution does not depend on the social objective. Consequently it is given by equations (7) with the implication that $c_D = c_L < c_H$ and $z_D = z_L < z_H$. Observe that to decentralize the full information solution, one simply needs lump-sum transfers from H and L to D and between H and L (the direction depending on the weights). The retirement choice does not have to be distorted. This is because the weighted objective function necessarily yields a Pareto-efficient solution so that the second welfare theorem applies.

3.2 Unobservable types: second-best solution

Let us now turn to the second-best problem, where types are not observable. With three types, we then have to add six incentive compatibility constraints.³ In particular, we have

$$u(c_L) - R_L(z_L) \geq u(c_D) - R_L(z_D) \quad (9)$$

for type L and

$$u(c_D) - R_D(z_D) \geq u(c_L) - R_D(z_L) \quad (10)$$

³Each type can potentially mimick either of the two other types.

for type D . Recalling that $R_L(z) \equiv R_D(z)$ and combining (10) and (9) yields not surprisingly:

$$u(c_L) - R_L(z_L) = u(c_D) - R_L(z_D). \quad (11)$$

In words, because D and L have the same preferences over c and z , their utility levels must be equalized at any incentive-compatible allocation. The simplest way to achieve (11) is of course to offer the same consumption bundle to the two types $c_L = c_D$ and consequently $z_L = z_D$. As will become clear below, this is effectively optimal even when it is not a priori imposed.⁴

When types D and L are pooled, the problem of the social planner is

$$\begin{aligned} \max \sum_i \varphi_i p_i [u(c_i) - R_i(z_i)] + \mu \sum_i p_i (wz_i - c_i) \\ + \lambda [u(c_H) - R_H(z_H) - u(c_L) + R_H(z_L)] \end{aligned}$$

where we use the subscript L for both L and D 's choices of z and c . The FOC with respect to z_H and c_H yield the traditional no distortion at the top: the marginal rate of substitution of type H continues to be given by (7). The FOC with respect to $z_L = z_D$ and $c_L = c_D$ yield:

$$\frac{R'_L(z_L)}{u'(c_L)} = w - \frac{\lambda}{(\varphi_L p_L + \varphi_D p_D - \lambda)} \frac{[R'_L(z_L) - R'_H(z_L)]}{u'(c_L)}. \quad (12)$$

where $(\varphi_L p_L + \varphi_D p_D - \lambda) > 0$ follows from the FOC with respect to $c_L = c_D$, while the single-crossing property (3) implies $R'_L(z_L) > R'_H(z_L)$. Consequently, we have a downward distortion and a positive marginal tax on $z_L = z_D$. We have a rather standard optimal tax problem. A distortion on z is here only justified by incentive arguments. Per our assumptions on preferences, a downward distortion is called for to relax an otherwise binding self-selection constraint.

3.3 Second-best with auditing

It is pretty clear that the lack of observability of the health status can be quite costly in terms of welfare. It is particularly detrimental to the disabled workers. Because they

⁴That is when any bundles (c_L, z_L) and (c_D, z_D) satisfying (11) are permitted. We do not prove this property here because the easiest proof makes use of the FOC derived in the next subsection; see Remark 1 in Appendix A1. Observe that such a pooling may also arise when D and L do not have identical preferences.

are pooled with the leisure-prone ones they are forced to work longer and receive less benefits than they would otherwise in a world of perfect observability.

To improve upon this second-best solution, one has to find ways to distinguish types L and D . A traditional way to achieve this is to use disability evaluations.⁵

To study the implication of such a policy, we adopt a very simple form of audit which perfectly reveals an individual's "health status".⁶ More precisely, the audit indicates whether or not an individual is of type D . In particular, it can be used to distinguish between D and L individuals. Observe that the audit also reveals mimicking between H and D , but *not* between H and L (who are both in good health). When a fraction π of type D 's reports are subject to testing, the corresponding audit cost is given by $k(p_D\pi)$. An individual who is caught cheating is subject to a sanction which reduces his utility to some exogenous and "low" utility level of \underline{u} . With this assumption, we have to modify the revenue constraints of the government and consider six incentive compatibility relations:

$$\sum_i p_i (wz_i - c_i) - k(p_D\pi) = 0$$

$$(LD) : u(c_L) - R_L(z_L) - (1 - \pi)(u(c_D) - R_L(z_D)) - \pi\underline{u} \geq 0$$

$$(DL) : u(c_D) - R_L(z_D) - u(c_L) + R_L(z_L) \geq 0$$

$$(HD) : u(c_H) - R_H(z_H) - (1 - \pi)(u(c_D) - R_H(z_D)) - \pi\underline{u} \geq 0$$

$$(DH) : u(c_D) - R_L(z_D) - u(c_H) + R_L(z_H)$$

$$(HL) : u(c_H) - R_H(z_H) - u(c_L) + R_H(z_L) \geq 0$$

$$(LH) : u(c_L) - R_L(z_L) - u(c_H) + R_L(z_H) \geq 0.$$

The properties of the solution will of course depend on which of the incentive constraints are binding at the optimum. With six incentive the number of potential com-

⁵An alternative way, considered by Marchand *et al.* (2003), is to observe the consumption vectors of both types which surely reveal their differences. One can on that basis self-select the two types with a solution that can be implemented with non-linear taxes on consumption and income.

⁶There is some literature on this. In general the tests are imperfect generating errors of types 1 and 2; see Diamond and Sheshinski (1995) and Parsons (1996). As long as the probability of audit is lower than one, our audit process implies three categories of individuals: able (or leisure-prone) individuals who are not audited and disabled individuals who are either audited or not. In Diamond and Sheshinski (1995) the two latter groups may contain able individuals because of type 1 and 2 errors in the audit process. This difficulty does not arise in our setting where audits are perfect.

binations of binding incentive constraint is quite large. To reduce the number of cases we shall first of all make some rather plausible assumptions. First, we assume that the weights are such that the “upwards” constraints (DH) and (LH) are not binding. With $\varphi_D = \varphi_H > \varphi_L$ it is clear that (LH) can potentially be binding when the weight on L becomes very small. We assume that this is not the case. Second, we shall restrict our attention to settings where (LD) is binding. This, in turn immediately implies that (DL) is not binding (as long as $\pi > 0$).

The Lagrangian expression associated with the problem of the social planner is then given by:

$$\begin{aligned}\mathcal{L}_1 = & \sum_i \varphi_i p_i [u(c_i) - R_i(z_i)] + \mu \left[\sum_i p_i (wz_i - c_i) - k(p_D \pi) \right] \\ & + \lambda_{LD} [u(c_L) - R_L(z_L) - (1 - \pi)(u(c_D) - R_L(z_D)) - \pi \underline{u}] \\ & + \lambda_{HL} [u(c_H) - R_H(z_H) - u(c_L) + R_H(z_L)] \\ & + \lambda_{HD} [u(c_H) - R_H(z_H) - (1 - \pi)(u(c_D) - R_H(z_D)) - \pi \underline{u}]\end{aligned}$$

The FOC are provided in Appendix A1; they can be rearranged as follows:

$$MRS_H = \frac{R'_H(z_H)}{u'(c_H)} = w \quad (13)$$

$$MRS_L = \frac{R'_L(z_L)}{u'(c_L)} = w - \frac{\lambda_{HL} [R'_L(z_L) - R'_H(z_L)]}{(\varphi_L p_L + \lambda_{LD} - \lambda_{HL}) u'(c_L)} \leq w \quad (14)$$

$$MRS_D = \frac{R'_D(z_D)}{u'(c_D)} = w - \frac{\lambda_{HD} (1 - \pi) [R'_L(z_D) - R'_H(z_D)]}{\varphi_D p_D - (\lambda_{LD} + \lambda_{HD}) (1 - \pi) u'(c_D)} \leq w \quad (15)$$

In the Appendix, we use these first-order-conditions to derive the following lemma which further reduces the number of regimes to be considered.

Lemma 1 *A solution with $\lambda_{HD} > 0$ and $\lambda_{HL} = 0$ is not possible.*

Lemma 1 along with the assumption made above imply that we are left with the following three regimes.

Regime	Binding incentive constraints
1	(HL),(HD),(LD)
2	(HL),(LD)
3	(LD)

The numerical illustrations provided in the next subsection show that each of these three regimes can effectively occur. From the first-order conditions we then obtain the following results for the different types of individuals.

Type H Whatever the regime we have from (13) that the *marginal* tax on z is zero for the type H individual. This is the usual no distortion at the top result which arises because no incentive constraint towards these individuals is binding.

Type L We obtain from (14) that this individual is subject to a downward distortion on their choice of retirement if (and only if) $\lambda_{HL} > 0$ that is in Regimes 1 and 2.

Type D Using (15) we see that there may or may not be a distortion in the choice of z for this type, depending on λ_{HD} . When $\lambda_{HD} > 0$, z_D is subject to a downward distortion which does not come as a surprise. This occurs in Regime 1 only. In the two other regimes where $\lambda_{HD} = 0$ individuals of type D are subject to no distortion. This may appear surprising at first because an incentive constraint towards D , namely (LD) is binding. However, since D and L have the same indifference curves, a distortion would not be an effective way to relax this incentive constraint.⁷

An interesting question which arises is whether the introduction of audit makes the lazy retire later than the disabled? The answer depends on the regime. In Regime 3, neither type D or L are subject to distortions. Since type D workers are on a higher indifference curve, we unambiguously have $z_L > z_D$. In Regime 2, there is an implicit tax on z_L but the disabled receive a higher transfer than the lazy. These transfers have an effect that fosters early retirement so that the comparison between z_L and z_D is ambiguous. In Regime 1 where the comparison between z_D and z_L seems ambiguous (both being distorted downwards), we can rank them unambiguously. To see this, we write (HL) , (LD) and (HD) as equalities and combine them, yielding:

$$R_L(z_L) - R_H(z_L) = (1 - \pi) [R_L(z_D) - R_H(z_D)]. \quad (16)$$

⁷To be more precise, the mimicked individual would be hurt as much as the mimicker. Consequently, a distortion cannot be welfare-improving.

The LHS of this expression is the utility differential between H and L at L 's consumption bundle. It gives the rent which must be conceded to H to prevent mimicking of L . The RHS is also a utility differential between H and L but at D 's consumption bundle. This shows the *extra* rent H must be granted compared to L at a solution where *both* types are prevented from mimicking D . This second rent is weighted by the probability of not being detected by the audit since types H and L are subject to audit when mimicking type D . Not surprisingly, when (HL) , (LD) and (HD) are binding the two rents in (16) are equalized. These two rents being increasing in z by convexity of $R(\cdot)$, one has $z_D > z_L$. Consequently, in Regime 1 we do have the result that disabled individuals retire later than the lazy ones.

All these results are valid for *any arbitrary* level of the audit probability ($0 < \pi < 1$).⁸ To determine the *optimal* level of π , we have to use the first-order condition (A7). It shows that the determination of π involves a tradeoff between costs ($\mu p_D k'$) and benefits that depend on the utility gap between truthtelling and being caught. Quite evidently the larger the audit costs, and the larger μ the lower will be π . Conversely, the lower the audit costs, the higher will be π . Note, however, that auditing with probability one ($\pi = 1$) is never optimal. This is a standard result in auditing models. It directly follows from (A7) by observing that when $\pi = 1$, (LD) and (HD) are satisfied with strict inequality consequently we would have $\lambda_{LD} = \lambda_{HD} = 0$ which implies that the derivative is negative. This makes sense, when incentive constraints are satisfied with strict inequality, they continue to be satisfied when audit probability (and hence audit cost) is slightly reduced.

Not much can be said analytically about the factors explaining the occurrence of one regime rather than another. However, it is clear that π , which in turn is determined by audit costs will play a crucial role. In particular, when π is high enough, Regime 1 cannot occur. To see this suppose that (LD) and (HL) bind. Using the expressions for the incentive constraints it follows that (HD) is not binding when

$$R_L(z_L) - R_H(z_L) - (1 - \pi)[R_L(z_D) - R_H(z_D)] > 0. \quad (17)$$

⁸Consequently, they also apply when the audit probability is exogenously given (rather than set at its optimal level).

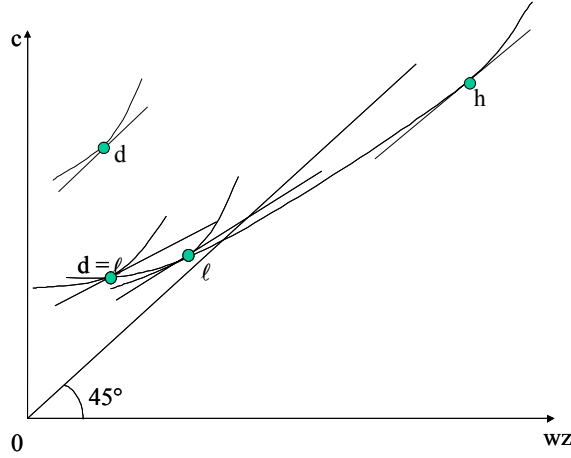


Figure 1: Solution without and with auditing.

which necessarily holds when π is close to one.

Figure 1 illustrates the above solution for Regime 2 as well as the solution without audit studied in Subsection 3.2. Without audit, individuals of type H consume the bundle h and the individuals D and L who cannot be distinguished consume $\ell = d$. With audit, we have ℓ and d for the lazy and the disabled respectively. Now the disabled could get a consumption level as high as that of the individuals h . However, while this would be true at the first-best, it is of course not desirable in a second-best setting.⁹ With $\lambda_{HD} = 0$ (Regime 2), individuals D do not face any distortion.¹⁰ The comparison between z_D and z_L is ambiguous. Type D individuals are on a higher indifference curve; this income effect decreases labor supply. Type L individuals are subject to a positive marginal tax which tends to decrease labor supply. Summing up, one can expect individuals of type D to work less (retire earlier) than individuals of type L if the income effect is sufficiently important (or if audit probability is large). The individuals L work more and consume also more than in the pooling equilibrium. We now turn to a numerical illustration.

⁹This follows directly from (A1) and (A3).

¹⁰Regime 1 yields a similar representation, except that D now faces a distortion.

3.4 Illustration

The functional form we use is given by:

$$u(c_i, z_i) = \ln c_i - \alpha_i \frac{z_i^2}{2},$$

with the following parameter values

	H	L	D
α_i	1	5	5
p_i	0.6	0.1	0.3
φ_i	1	0.3	1
w_i	100	100	100

We consider three audit technologies depending on their cost: audit 1 is relatively expensive ($k(\pi) = 2000\pi^2$) and implies $\pi = 0.06$, $\lambda_{HL} > 0$ and $\lambda_{LD} > 0$ and $\lambda_{HD} > 0$ (Regime 1); audit 2 is cheaper ($k(\pi) = 1000\pi^2$) and implies $\pi = 0.10$ with $\lambda_{HL} > 0$ and $\lambda_{LD} > 0$ (Regime 2). Finally audit 3 is the cheapest ($k(\pi) = 100\pi^2$) and implies $\pi = 0.49$ (and Regime 3). In all cases, $\underline{u} = 0$. The results are given in Table 1.

Types	First-best			Pooling		
	D	L	H	D	L	H
c	88.39	26.51	88.39	54	54	92.48
z	0.22	0.75	1.13	0.30	0.30	1.08
u	4.35	1.85	3.84	3.75	3.75	3.94
$T'(z)$	0	0	0	0.17	0.17	0

Types	Audit 1			Audit 2			Audit 3		
	D	L	H	D	L	H	D	L	H
c	65.29	49.72	89.32	70.74	48.23	87.94	81.42	46.59	84.95
z	0.29	0.28	1.11	0.28	0.30	1.13	0.24	0.43	1.17
u	3.95	3.70	3.86	4.05	3.64	3.83	4.24	3.38	3.74
$T'(z)$	0.03	0.28	0	0	0.26	0	0	0	0

Table 1: Numerical example, Paretian objective

First, observe that starting from the no auditing pooling solution, introducing audit 1 leads to Regime 1 where all three incentive compatibility constraints are binding. Turning to a less expensive technology (and thus to a higher level of audit) implies a

regime where (HD) no longer binds. This illustrates the argument we made on the basis of (17). For even smaller audit costs we obtain Regime 3. Observe, however, that if the audit technology would be completely costless, we would have either only (HL) binding if φ_L is relatively high or the first best if φ_L is relatively low¹¹.

Second, we observe that in all the second-best solutions the consumption of D is kept below that of H whereas in the first-best the two types have the same consumption. Not surprisingly, the difference between c_H and c_D decreases as the audit probability increases. Furthermore, in the pooling solution we have $c_D = c_L$ while with auditing $c_D > c_L$ holds; as expected the difference between c_D and c_L increases as audit become cheaper and more intense and thus allows for a better screening. Finally, in Regime 1 D retires later than L which is in line with the analytical result obtained above. In the other two regimes, however, we have $z_D < z_L$, even though L faces a positive marginal tax in Regime 2 while D 's marginal tax is zero. This illustrates the idea that the income effects (transfer to D) can dominate the substitution effects (marginal distortions).

4 Alternative specification

We now consider an alternative specification wherein the social planner paternalistically attributes to the lazy workers a disutility R_L rather than R_H . In other words, he wants them to work more and not to consume less. Once again, we start by considering the full information solution.

4.1 Full information solution and decentralization

With the paternalistic approach the full information problem is given by:

$$\max \sum_i p_i u(c_i) - p_H R_H(z_H) - p_L R_H(z_L) - p_D R_L(z_D) + \mu \sum_i p_i (c_i - w z_i).$$

The solution is easily obtained. Both healthy types (namely H and L) are assigned the same retirement age, the disabled individuals retire earlier and all individuals have the

¹¹If the first best allocation described in section 3.1 respects (HL) then the first best allocation is implementable with audits. If it is not (which is the case if φ_L is high enough) then (HL) would bind with free audit.

same consumption level:

$$z_D < z_L = z_H \quad \text{and} \quad c_D = c_L = c_H.$$

The *laissez-faire* solution does not depend on the social objective and is thus the same as in Subsection 3.1. Consequently it is given by equations (7) with the implication that $c_D = c_L < c_H$ and $z_D = z_L < z_H$. Decentralizing the first-best, however, is now not as simple as in the non-paternalistic specification. Lump sum transfers are not sufficient; one also needs a subsidy (negative marginal tax), on the labor supply of type L individuals. To see this, note that the decentralized choice of z_L is determined by:

$$(w - T'(z_L))u'(c_L) - R'_L(z_L) = 0 \tag{18}$$

which coincides with the socially optimal one

$$w u'(c_L) - R'_H(z_L) = 0. \tag{19}$$

when the marginal tax rate is given by

$$T'(z_L) = -\frac{[R'_L(z_L) - R'_H(z_L)]}{u'(c_L)} < 0, \tag{20}$$

so that we effectively have a subsidy. In words, the rate of subsidy is equal to the difference between “private” and the “social” marginal valuation for z_L (evaluated at the optimal allocation). Paternalism introduces a wedge between private and social valuation of type L ’s labor supply and the decentralization of the optimum requires a “Pigouvian” subsidy.

4.2 Unobservable types: second-best solution

Let us now turn to the second-best problem, where types are not observable. With three types, we then have to consider six incentive constraints. As in Section 3, we start with the no-audit case, for which one can show that only (HL) , (HD) and (LD) can possibly be binding.¹² Clearly, when (HD) is binding together with (HL) and (LD) ,

¹²In other words the upwards constraint (LH) and (DH) cannot be binding. While this property is not surprising, its proof is somewhat tedious. It is omitted here but is available from the authors upon request.

one has a pooling optimum. The program is now:

$$\begin{aligned}
& \max \sum p_i u(c_i) - p_H R_H(z_H) - p_L R_H(z_L) - p_D R_L(z_L) \\
& + \mu \sum p_i (wz_i - c_i) + \lambda_{HL} [u(c_H) - R_H(z_H) - u(c_L) + R_H(z_L)] \\
& + \lambda_{HD} [u(c_H) - R_H(z_H) - u(c_D) + R_H(z_D)] \\
& + \lambda_{LD} [u(c_L) - R_L(z_L) - u(c_D) + R_L(z_D)]
\end{aligned}$$

The first-order conditions with respect to z_H and c_H yield the traditional no distortion at the top result: the marginal rate of substitution of type H continues to be given by (7). The FOC with respect to z_L , z_D , c_L and c_D are:

$$(p_L - \lambda_{HL} + \lambda_{LD}) u'(c_L) + \mu p_L = 0 \quad (21)$$

$$(p_D - \lambda_{HD} - \lambda_{LD}) u'(c_D) + \mu p_D = 0 \quad (22)$$

$$-p_L R'_H(z_L) - \mu p_L w + \lambda_{HL} R'_H(z_L) - \lambda_{LD} R'_L(z_L) = 0 \quad (23)$$

$$-p_D R'_L(z_D) - \mu p_D w + \lambda_{HD} R'_H(z_D) + \lambda_{LD} R'_L(z_D) = 0 \quad (24)$$

This yields:

$$MRS_L = \frac{R'_L(z_L)}{u'(c_L)} = w + \left[\frac{p_L - \lambda_{HL}}{p_L - \lambda_{HL} + \lambda_{LD}} \right] \left[\frac{R'_L(z_L) - R'_H(z_L)}{u'(c_L)} \right], \quad (25)$$

where $p_L - \lambda_{HL} + \lambda_{LD} > 0$ from (21) while the sign of $(p_L - \lambda)$ appears to be ambiguous. Furthermore we have

$$MRS_D = \frac{R'_D(z_D)}{u'(c_D)} = w - \frac{\lambda_{HD}}{(p_D - \lambda_{LD} - \lambda_{HD})} \left[\frac{R'_L(z_D) - R'_H(z_D)}{u'(c_D)} \right] \quad (26)$$

where $(p_D - \lambda_{LD} - \lambda_{HD}) > 0$ from (22) so that we necessarily have a downward distortion for type D 's choice of z . The new feature which arises here is that D and L are no longer necessarily pooled even when there is no audit. More precisely, we have

Lemma 2 *The optimum pools D and L if and only if $R'_L(z_L)/u'(c_L) < w$ that is iff $p_L < \lambda_{HL}$.*

To interpret this result, consider equation (25). The distortion on z_L is of ambiguous sign. When $p_L > \lambda_{HL}$, the RHS of (25) is larger than w so that we have an upward

distortion on labor supply (a negative marginal tax rate). Conversely, $p_L < \lambda_{HL}$ yields a downward distortion and thus a positive marginal tax. To understand the role played by this term, note that the distortion in type L 's labor supply is determined by balancing two conflicting objectives. First, there is the standard (Mirrlees-type) incentive effect which calls for a downward distortion (positive marginal tax) on L 's labor supply in order to relax the incentive constraint from type H to type L . The significance of this effect depends on λ_{HL} , the Lagrange multiplier of this incentive constraint. Second, there is a “paternalistic” effect, which calls for an increase in type L 's labor supply compared to the laissez-faire level. Recall, that social disutility of their labor is smaller than their private disutility. This pleads in favor of an upward distortion (a negative marginal tax). This effect is the more significant, the larger the proportion of group L in the economy. When p_L is large enough, the paternalistic consideration is dominant so that there is a marginal subsidy on z_L . In this case, there is no reason to pool individuals L and D because there is no need of distortion on z_D . However, when p_L is small, incentive considerations dominate and a marginal tax is called for on z_L . In this case, one necessarily pools individuals L and D in order to prevent individuals H from mimicking individuals D .

To sum up, when p_L is large enough, the optimum entails separation between type D and L individuals. In this case, there is no distortion on z_D and a marginal subsidy on z_L . When p_L is small, one has a pooling optimum with a marginal tax on z_L and z_D . Note finally, that in the separating optimum, one necessarily always has $z_L > z_D$.

4.3 Second-best with audit in the paternalistic case

Introducing audit as in Section 3, while continuing to assume that the upwards constraints (LH) and (DH) are not binding, the social planner problem can now be written

as the maximization of:

$$\begin{aligned}
\mathcal{L}_2 = & \sum_{i=1}^3 p_i u(c_i) - p_H R_H(z_H) - p_L R_H(z_L) - p_D R_L(z_D) \\
& + \mu \left[\sum_{i=1}^3 p_i (w z_i - c_i) - p_D k \pi \right] \\
& + \lambda_{LD} [u(c_L) - R_L(z_L) - (1 - \pi)(u(c_D) - R_L(z_D)) - \pi \underline{u}] \\
& + \lambda_{HL} [u(c_H) - R_H(z_H) - u(c_L) + R_H(z_L)] \\
& + \lambda_{HD} [u(c_H) - R_H(z_H) - (1 - \pi)(u(c_D) - R_H(z_D)) - \pi \underline{u}] \\
& + \lambda_{DL} [u(c_D) - R_D(z_D) - u(c_L) + R_D(z_L)]
\end{aligned}$$

The FOC are stated Appendix A3. Combining and rearranging these conditions yield the following expressions for the marginal rates of substitution of the different types:

$$\frac{R'_H(z_H)}{u'(c_H)} = w. \quad (27)$$

$$\frac{R'_L(z_L)}{u'(c_L)} = w + \frac{(p_L - \lambda_{HL}) [R'_L(z_L) - R'_H(z_L)]}{(p_L + \lambda_{LD} - \lambda_{DL} - \lambda_{HL}) u'(c_L)} \begin{matrix} \leq \\ \geq \end{matrix} w \quad (28)$$

$$\frac{R'_L(z_D)}{u'(c_D)} = w - \frac{\lambda_{HD} (1 - \pi) [R'_L(z_D) - R'_H(z_D)]}{(p_D - (\lambda_{LD} + \lambda_{HD}) (1 - \pi) + \lambda_{DL}) u'(c_D)} \leq w. \quad (29)$$

Before commenting on these expressions, it is necessary to determine the possible regimes. To reduce the number of cases, we can first note that Lemma 1 continues to hold in the alternative specification.¹³ Next, Appendix A5 provides a sequence of additional Lemmas which show that in the paternalistic case, only two regimes are possible which are given by:

Regime	Binding incentive constraints
1	$(HL), (HD), (LD)$
2	$(HL), (LD)$

Consequently, it turns out that Regime 3, where only (LD) was binding cannot arise here.

Let us now turn to the interpretation of conditions (27)–(29) while taking into account that in both possible regimes (DL) is not binding so that we have $\lambda_{DL} = 0$.

¹³The proof in Appendix A2 does not rely on the objective function of the government.

Equation (27) is the usual no distortion at the top property. Turning to (28) observe that $(p_L + \lambda_{LD} - \lambda_{HL}) > 0$ from (A12), while $(p_L - \lambda_{HL})$ is ambiguous. As in the preceding section, when $p_L > \lambda_{HL}$, paternalistic considerations are dominant and call for a marginal subsidy on z_L . When $p_L < \lambda_{HL}$, incentive considerations dominate and a positive marginal tax on z_L is optimal.

For the interpretation of (29), on the other hand, we have to distinguish Regime 2 with $\lambda_{HD} = 0$ from Regime 1 with $\lambda_{HD} > 0$. The results are exactly like in Section 3. In Regime 2, $\lambda_{HD} = 0$, there is no distortion in the choice of retirement for the truly disabled. The mimicking individual L has the same indifference curves as individual D . Consequently, there is no way to create a distortion which hurts the mimicker more than the mimicked. Put differently a distortion is not an effective way to relax the incentive constraint. Consequently, the retirement choice of type D is treated *at the margin* exactly like that of the type H individual (zero marginal tax). In Regime 1, on the other hand, we have $\lambda_{HD} > 0$, and there is a distortion towards earlier retirement. This distortion relaxes the incentive constraint (HD). As in Section 3, this distortion is likely to be effective for a relatively low probability of auditing.

The comparison between z_L and z_D is the same as in Section 3. Note however that Regime 1 cannot occur when there is a marginal subsidy on z_L . To see this, recall that Regime 1 implies a distortion on z_D and $z_L < z_D$. Since D is on a higher indifference curve, this is incompatible with a marginal subsidy on z_L . We summarize this in the following lemma:

Lemma 3 *When $p_L > \lambda_{HL}$, Regime 1 cannot occur.*

Figure 2 provides a graphical representation of the problem with auditing. On the figure we have three points in the (y, c) space. The consumption bundle for individuals of type H is represented by h , where the slope of the indifference curve is 1. The distance between h and the (no tax) budget line $c = wz$ is the transfer individuals of type H have to pay. Point ℓ is the solution for individuals of type L with slope different from 1. We have represented the case where the slope is higher than 1, but it could have been less. Here individuals L are subject to a marginal subsidy and receive a transfer. Note

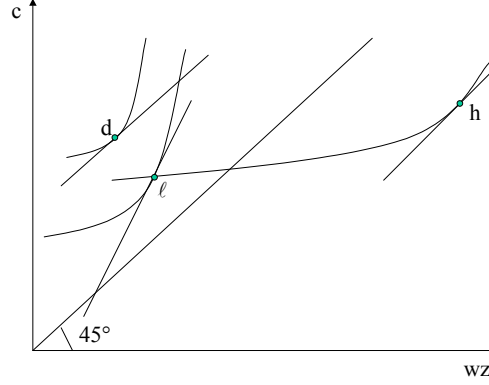


Figure 2: Paternalistic objective: solution with auditing

the binding incentive constraint from H to L and the relative slope at ℓ of these two types indifference curves. Finally, d is the solution for individuals of type D . The slope is 1 (no distortion as for individuals H); they receive a larger transfer than individuals L . They also work less (d is to the left of ℓ). Note however that with a positive marginal tax on L and a weak income effect, it is not impossible to have $z_D > z_L$ in the second best (recall that it is always the case in Regime 1). Finally, the indifference curve going through ℓ is a linear combination of that going through d and one yielding a level of utility \underline{u} .

Equation (A16) determines the optimal value of π which depends on both k , the cost of auditing, and \underline{u} , the penalty. Not surprisingly, the optimal audit probability (for type D reports) is determined by the tradeoff between its negative effect on public expenditures ($-\mu p_D k$) and its positive effect via the welfare gain stemming from less stringent IC constraints.

Types	First best			No audit		
	D	L	H	D	L	H
c	86.74	86.74	86.74	57.72	57.72	94.04
z	0.28	1.15	1.15	0.39	0.39	1.06
u	4.29	1.80	3.79	3.74	3.74	3.97
$T'(z)$	0	-3	0	0.09	0.09	0

Types	Audit 1			Audit 2		
	$\pi = 0.04$, Regime 1			$\pi = 0.08$, Regime 2		
Types	D	L	H	D	L	H
	D	L	H	D	L	H
c	65.35	54.05	91.55	68.06	53.70	90.59
z	0.38	0.37	1.09	0.36	0.41	1.10
u	3.88	3.71	3.92	3.95	3.63	3.89
$T'(z)$	0.003	0.19	0	0	0.11	0

Table 3: Numerical example: Paternalistic objective, Scenario 1.

4.4 Illustration: paternalistic objective

We use the same utility function as in Section 3 and continue to assume $\underline{u} = 0$. However in order to illustrate the pooling and the separating optima, we will consider the two following scenarios for the parameter values:

	Scenario1			Scenario2		
	H	L	D	H	L	D
α_i	1	4	4	1	4	4
p_i	0.6	0.07	0.33	0.6	0.2	0.2
w_i	100	100	100	100	100	100

These two scenarios differ in the proportion of type L and D in the economy. In Scenario 1, there is a small proportion of type L individuals while this proportion is larger in Scenario 2. In addition, we consider two audit technologies depending on their cost: Audit 1 is relatively expensive ($k(\pi) = 2000\pi^2$) while Audit 2 is cheaper ($k(\pi) = 1000\pi^2$). The results are given in Table 3 (Scenario 1) and Table 4 (Scenario 2).

Not surprisingly, Audit 1 implies lower audit probabilities than Audit 2. In scenario 1, p_L is relatively low so that without audit, a pooling optimum will occur with a positive marginal tax on z_L and z_D . As in the case of the Paretian objective, as audits increase, one goes from Regime 1 to 2. In Regime 1, lazy individuals will work less than the

FirstBest				No audit		
Types	D	L	H	D	L	H
c	92.19	92.19	92.19	52.20	70.07	97.8
z	0.27	1.08	1.08	0.47	0.61	1.02
u	4.37	2.17	3.93	3.49	3.49	4.06
$T'(z)$	0	-3	0	0	-0.71	0

Audit 1				Audit 2		
$\pi = 0.12$, Regime 2				$\pi = 0.20$, Regime 2		
Types	D	L	H	D	L	H
c	60.54	70	95.67	67.20	71.50	94.42
z	0.41	0.68	1.04	0.37	0.75	1.05
u	3.76	3.31	4.01	3.93	3.13	3.98
$T'(z)$	0	-0.91	0	0	-1.15	0

Table 4: Numerical example: Paternalistic objective, Scenario 2.

disabled while in Regime 2 they turn to work more.

In scenario 2, the optimum without audit involves separation between individuals D and L . This is because p_L is now relatively high so that paternalistic considerations dominate. In this case, the second best without and with audit involves no distortion on z_D while a marginal subsidy is called for on z_L .

5 Conclusion

We have studied the design of retirement cum disability benefits in a three type setting. Healthy individuals have a low disutility of labor (continued activity). The two other types have identical preferences with a high disutility of labor. However, they differ in their (unobservable) health status. Type L individuals have a high disutility for labor because they are leisure prone. Type D individuals, on the other hand, have a high disutility because they are disabled. We start from the premise that policy makers are not prepared to treat types D and L alike. Helping type D individuals and allowing them to retire early with “generous” benefits is considered as legitimate. However, society does not want to extend such a generous treatment to the L type individuals. Formally, this is introduced by considering either a (Paretian) social welfare function

which puts less weight on the type L individuals, or by using a paternalistic social welfare function which does not fully account for the type L individual's disutility of labor.

In either setting, a first-best solution would imply earlier retirement for D than for L ($z_D < z_L$). However, when the health status is not observable, it may or may not be possible to separate the two types out, so that the second-best may imply $z_D = z_L$. In that case, the effectiveness of disability benefits is severely undermined by the presence of the L type and the possibilities to help the disabled are limited. The situation can be improved if costly audits (test for disability) become available. However, the second-best with audit does not necessarily imply $z_D < z_L$; quite surprisingly, we may even end up with a situation where the disabled retire later than the lazy ($z_D > z_L$). But their consumption or rather their disability benefits are higher than those of workers of type L . Another interesting finding of our analysis is that the optimum often implies distorted retirement decision at least for types L and D . However, there are situation in which there are no distortion, even in the second best, or where the distortions affect only one of the types. We show that this depends on the pattern of binding incentive constraints which in turns depends on how costly the audit is.

To put these results in perspective, it is interesting to consider the policy implications of the NBER and OECD studies explaining early retirement by a too extensive use of social insurance programs is that one should make the overall system more neutral towards retirement decisions. Demographic aging makes such a reform extremely pressing. The problem with such a linear reform is that it is clearly legitimate as it regards workers who are not disabled nor involuntarily unemployed. It makes economic sense to decrease the generosity of social insurance programs for those who "abuse" the system. However, for those who are truly disabled or unemployed, such a reform necessarily implies lower benefits and this is questionable. From a public finance viewpoint, the reform works; from a social welfare viewpoint, it is likely to be catastrophic.

What we show in this paper is that to avoid such an outcome one should introduce or strengthen the audit and control techniques allowing to sort out leisure-prone workers and true disabled individuals. So doing one can at the same time keep the disability

benefits at a sufficient high level while inducing the non-disabled to work longer. We have focused on the disability insurance. The same approach should be used to deal with unemployment insurance.

References

- [1] Bound, J., (1989), The health and earnings of rejected disability insurance applicants, *American Economic Review*, 79, 482-503.
- [2] Crawford, V.P. and D.M. Lilien, (1981), Social security and the retirement decision, *Quarterly Journal of Economics*, 95, 505-529.
- [3] Cremer, H., J-M. Lozachmeur and P. Pestieau, (2004), Social security, retirement age and optimal income taxation, *Journal of Public Economics*, forthcoming.
- [4] Desmet, R. and A. Jousten, (2003), The decision to retire: individual heterogeneity and actuarial neutrality, unpublished.
- [5] Desmet, R., A. Jousten, S. Perelman and P. Pestieau, (2003), Micro-simulation of social security reforms in Belgium, IZA Discussion Paper #735.
- [6] Diamond, P. and J. Mirrlees, (1986), Payroll tax financed social insurance with variable retirement, *Scandinavian Journal of Economics*, 88, 25-50.
- [7] Diamond, P. and E. Sheshinski, (1995), Economic aspects of optimal disability benefits, *Journal of Public Economics*, 1-24.
- [8] Gruber, J. and D. Wise, (2004), *Social Security Programs and Retirement around the World: Micro-Estimations*, The Chicago University Press, Chicago, forthcoming.
- [9] Gruber, J. and D. Wise, (1999), *Social Security and Retirement around the World*, The Chicago University Press, Chicago.
- [10] Gruber, G., (2000), Disability insurance benefits and labor supply, *Journal of Political Economy*, 108, 1162-1183.

- [11] Marchand, M., P. Pestieau and M. Racionero, (2003), Optimal redistribution when different workers are indistinguishable, *Canadian Journal of Economics*, forthcoming.
- [12] Parsons, D., (1980), The decline of male labor force participation, *Journal of Political Economy*, 88, 117-134.
- [13] Parsons, D., (1991), The health and earnings of rejected disability insurance applicants: comment, *American Economic Review*, 81, 1419-1426.
- [14] Parsons, D.O (1996), Imperfect 'tagging' in social insurance programs, *Journal of Public Economics*, 62, 183-207.
- [15] Roemer, J., (1998), *Equality of Opportunity*, Harvard University Press, Cambridge.

Appendix

A1 First-order conditions for \mathcal{L}_1

Differentiating \mathcal{L}_1 yields the following FOC:

$$\frac{\partial \mathcal{L}_1}{\partial c_H} = \varphi_H p_H u'(c_H) - p_H \mu + (\lambda_{HL} + \lambda_{HD}) u'(c_H) = 0 \quad (\text{A1})$$

$$\frac{\partial \mathcal{L}_1}{\partial z_H} = -\varphi_H p_H R'_H(z_H) + p_H \mu w - (\lambda_{HL} + \lambda_{HD}) R'_H(z_H) = 0 \quad (\text{A2})$$

$$\frac{\partial \mathcal{L}_1}{\partial c_L} = \varphi_L p_L u'(c_L) - p_L \mu + (\lambda_{LD} - \lambda_{HL}) u'(c_L) = 0 \quad (\text{A3})$$

$$\frac{\partial \mathcal{L}_1}{\partial z_L} = -\varphi_L p_L R'_L(z_L) + p_L \mu w - \lambda_{LD} R'_L(z_L) + \lambda_{HL} R'_H(z_L) = 0 \quad (\text{A4})$$

$$\frac{\partial \mathcal{L}_1}{\partial c_D} = \varphi_D p_D u'(c_D) - p_D \mu - (\lambda_{LD} + \lambda_{HD}) (1 - \pi) u'(c_D) = 0 \quad (\text{A5})$$

$$\begin{aligned} \frac{\partial \mathcal{L}_1}{\partial z_D} &= -\varphi_D p_D R'_L(z_D) + p_D \mu w + \lambda_{LD} (1 - \pi) R'_L(z_D) \\ &\quad + \lambda_{HD} (1 - \pi) R'_H(z_D) = 0 \end{aligned} \quad (\text{A6})$$

$$\frac{\partial \mathcal{L}_1}{\partial \pi} = -\mu p_D k' + \lambda_{LD} [u(c_D) - R_L(z_D) - \mathfrak{u}] + \lambda_{HD} [u(c_D) - R_H(z_D) - \mathfrak{u}] \leq 0. \quad (\text{A7})$$

Remark 1 *These first-order condition also apply to the case where $\pi = 0$, that is if there is no audit, but if we do not a priori impose pooling between L and D . They can be used to show that with $\pi = 0$, pooling is unavoidable. To see this recall that without audit, L and D are on the same indifference curve. A simple graphical argument shows that we have $\lambda_{HL} > 0$ and $\lambda_{HD} = 0$ when $z_L > z_D$ while we have $\lambda_{HD} > 0$ and $\lambda_{HL} = 0$ when $z_D > z_L$. However, when $\lambda_{HL} > 0$ and $\lambda_{HD} = 0$ we have from (14)–(15) that z_L is distorted downwards, while z_D is not distorted. But with both types on the same indifference curve, this implies $z_D > z_L$ and we have a contradiction. The case where $\lambda_{HD} > 0$ and $\lambda_{HL} = 0$ can be dealt with along the same lines.*

A2 Proof of Lemma 1

Proof. We prove this by contradiction. Assume

$$\lambda_{HD} > 0 \quad \text{and} \quad \lambda_{HL} = 0. \quad (\text{A8})$$

We then have

$$\begin{aligned}(LD) : u(c_L) - R_L(z_L) - (1 - \pi)(u(c_D) - R_L(z_D)) - \pi \underline{u} &\geq 0 \\(HD) : u(c_H) - R_H(z_H) - (1 - \pi)(u(c_D) - R_H(z_D)) - \pi \underline{u} &= 0.\end{aligned}$$

Using (HD) , (HL) holds with strict inequality if:

$$(1 - \pi)(u(c_D) - R_H(z_D)) + \pi \underline{u} > u(c_L) - R_H(z_L) \quad (A9)$$

Using (LD) , a necessary condition for $(A9)$ to hold is:

$$R_L(z_L) - R_H(z_L) < (1 - \pi)[R_L(z_D) - R_H(z_D)]$$

Because $R'_L(z) > R'_H(z)$, this requires $z_D > z_L$.

However, under conditions (A8), equations (28) and (29) imply $z_D < z_L$; this is because there is a marginal subsidy on z_L while there is a downward distortion for D and because $u(c_D) - R_L(c_D) \geq u(c_L) - R_L(c_L)$ (from incentive constraint (DL)) so that D is on a higher (or the same) indifference curve than L which implies that the income effect reinforces the substitution effect (assuming that leisure is normal). To sum up, we obtain a contradiction. ■

A3 First-order conditions for \mathcal{L}_2

Differentiating \mathcal{L}_2 yields the following FOC:

$$\frac{\partial \mathcal{L}_2}{\partial c_H} = p_H u'(c_H) - p_H \mu + (\lambda_{HL} + \lambda_{HD}) u'(c_H) = 0, \quad (\text{A10})$$

$$\frac{\partial \mathcal{L}_2}{\partial z_H} = -p_H R'_H(z_H) + p_H \mu w - (\lambda_{HL} + \lambda_{HD}) R'_H(z_H) = 0, \quad (\text{A11})$$

$$\frac{\partial \mathcal{L}_2}{\partial c_L} = p_L u'(c_L) - p_L \mu + (\lambda_{LD} - \lambda_{HL} - \lambda_{DL}) u'(c_L) = 0, \quad (\text{A12})$$

$$\frac{\partial \mathcal{L}_2}{\partial z_L} = -p_L R'_H(z_L) + p_L \mu w - (\lambda_{LD} - \lambda_{DL}) R'_L(z_L) + \lambda_{HL} R'_H(z_L) = 0 \quad (\text{A13})$$

$$\frac{\partial \mathcal{L}_2}{\partial c_D} = p_D u'(c_D) - p_D \mu + \lambda_{DL} u'(c_D) - (\lambda_{LD} + \lambda_{HD}) (1 - \pi) u'(c_D) = 0 \quad (\text{A14})$$

$$\begin{aligned} \frac{\partial \mathcal{L}_2}{\partial z_D} &= -p_D R'_L(z_D) + p_D \mu w - \lambda_{DL} R'_L(z_D) \\ &\quad + (\lambda_{LD} R'_L(z_D) + \lambda_{HD} R'_H(z_D)) (1 - \pi) = 0, \end{aligned} \quad (\text{A15})$$

$$\begin{aligned} \frac{\partial \mathcal{L}_2}{\partial \pi} &= -\mu p_D k + \lambda_{LD} (u(c_D) - R_L(z_D) - \underline{u}) \\ &\quad + \lambda_{HD} (u(c_D) - R_H(z_D) - \underline{u}) \leq 0. \end{aligned} \quad (\text{A16})$$

A4 Proof of Lemma 2

Proof. We proof this by contradiction. Point (i) contradicts $\frac{R'_L(z_L)}{u'(c_L)} < w$ with L and D 's separation. Point (ii) contradicts pooling D and L with $\frac{R'_L(z_L)}{u'(c_L)} > w$.

(i) Suppose $\frac{R'_L(z_L)}{u'(c_L)} < w$ and that we separate individuals L and D . It necessarily implies $\lambda_{HD} = 0$ otherwise the use of (HL) , (LD) and (HD) implies pooling. In this case, one has $\frac{R'_L(z_D)}{u'(c_D)} = w$ by (26). Thus, since (HL) and (LD) are binding, one necessarily has $u(c_H) - R_H(z_H) < u(c_D) - R_H(z_D)$ which contradicts (HD) .

(ii) Suppose we pool D and L with $\frac{R'_L(z_L)}{u'(c_L)} > w$. Then making a change $dc_D = \frac{R'_L(z_D)}{u'(c_D)} dz_D$ has no (first order) effect on individual D 's utility, (LD) and (HL) , and relaxes (HD) and the budget constraint. This contradicts the fact that we have an optimum. ■

A5 Regimes in the paternalistic case

Lemma 4 *If $\lambda_{DL} > 0$, then $\lambda_{HD} > 0$, $\lambda_{HL} > 0$ and $\lambda_{LD} = 0$.*

Proof. Suppose $\lambda_{DL} > 0$. By construction (LD) does not bind so that $\lambda_{LD} = 0$. Further, one necessarily needs $\lambda_{HD} > 0$ otherwise $\frac{\partial \mathcal{L}}{\partial \pi} < 0$ which contradicts $\pi > 0$. Because $\lambda_{HD} > 0$, one has $\lambda_{HL} > 0$ by Lemma 1. ■

Lemma 5 *A solution with $\lambda_{HD} > 0$, $\lambda_{HL} > 0$ and $\lambda_{LD} = 0$ is not possible.*

Proof. We prove this by contradiction. Assume

$$\lambda_{HD} > 0, \lambda_{HL} > 0 \quad \text{and} \quad \lambda_{LD} = 0. \quad (\text{A17})$$

We then have

$$(HL) : u(c_H) - R_H(z_H) - u(c_L) + R_H(z_L) = 0$$

$$(HD) : u(c_H) - R_H(z_H) - (1 - \pi)(u(c_D) - R_H(z_D)) - \pi \underline{u} = 0.$$

Using (HL) and (HD) , (LD) holds with strict inequality if

$$R_L(z_L) - R_H(z_L) < (1 - \pi)[R_L(z_D) - R_H(z_D)]$$

Because $R'_L(z) > R'_H(z)$, this requires $z_D > z_L$. Because $u(c_D) - R_L(z_D) \geq u(c_L) - R_L(z_L)$, $z_D > z_L$ implies $\frac{R'_L(z_D)}{u'(c_D)} > \frac{R'_L(z_L)}{u'(c_L)}$. Thus, since $\frac{R'_L(z_D)}{u'(c_D)} < w$ by (29), one necessarily has $\lambda_{HL} > p_L$. It implies $\frac{\partial \mathcal{L}_2}{\partial c_L} < 0$ since $\lambda_{LD} = 0$. This contradicts the fact that we have an interior optimum. ■

Lemma 6 *A solution with $\lambda_{DL} > 0$ is impossible.*

Proof. The result follows directly from the combination of Lemma 4 and 3. ■

Lemma 7 *A solution with $\lambda_{LD} > 0$ and $\lambda_{HD} = \lambda_{HL} = 0$ is not possible.*

Proof. We prove this by contradiction. Assume

$$\lambda_{HD} = \lambda_{HL} = 0 \quad \text{and} \quad \lambda_{LD} > 0. \quad (\text{A18})$$

First note that because $\lambda_{LD} > 0$, one has $\lambda_{DL} = 0$. Using (A10) and (A12), we have

$$\frac{u'(c_H)}{u'(c_L)} = 1 + \frac{\lambda_{LD}}{p_L} (1 - \pi)$$

which implies $c_L > c_H$.

By the same way, combining (A11) and (A13), we obtain

$$\frac{R'_H(z_H)}{R'_H(z_L)} = \frac{\mu w}{\mu w - \lambda_{LD} R'_L(z_L)}$$

which implies $z_L < z_H$. This contradicts (HL). ■

Lemma 8 *A solution with $\lambda_{HL} > 0$ and $\lambda_{HD} = \lambda_{LD} = 0$ is not possible.*

Proof. If it were the case, we would have $\frac{\partial \mathcal{L}_2}{\partial \pi} < 0$ which would contradict $\pi > 0$. ■